

Exercise 5

Lecture 8: Time Series Data

Scalable Data Engineering

5.1 Are the following statements true or false? Give reasons for your answers.

- a) Classical decomposition splits every value of a time series into a trend, season, and residuals.
- b) The season mask of classical decomposition can vary across a time series.
- c) Symbolic aggregate approximation represents segments of a time series by its discretized mean value.
- d) Data-mining tasks such as, e.g., classification and clustering, can be efficiently applied to the raw value representation of time series.
- e) For forecasting, time series only have to consist of time stamps and measured values.
- f) The historical measured values of a time series are always enough to accurately forecast a time series.
- g) The AIC uses the model complexity as well as the model fit to rate a model.

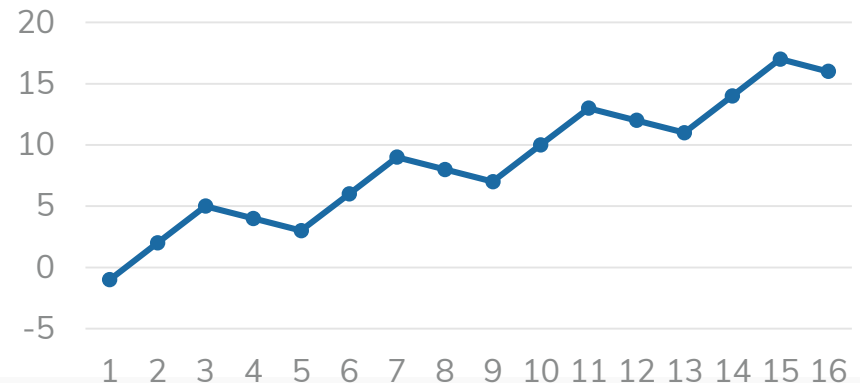
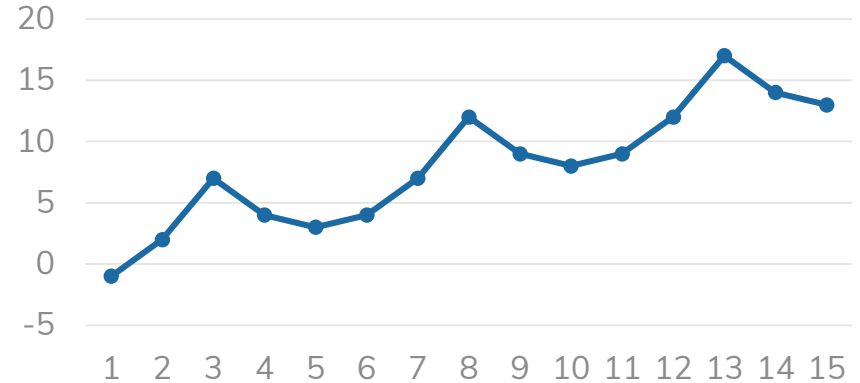
Classical Decomposition

5.2. a) Decompose the following time series into trend, season, and residuals.

- [-1, 2, 7, 4, 3, 4, 7, 12, 9, 8, 9, 12, 17, 14, 13]
- Visualize the time series.
- How long is one season?
- Detrend with a moving average filter of the season length.
- Retrieve the season mask by averaging the detrended time series values.

5.2. b) Which problem occurs if the following time series needs to be decomposed?

- [-1, 2, 5, 4, 3, 6, 9, 8, 7, 10, 13, 12, 11, 14, 17, 16]



5.3 Compare the forecast models on the given time series based on their AIC.

- [4, 7, 12, 9, 8, 9, 12, 17, 14, 13]
- Which of the following models is the best one?
- AR(1): $\hat{x}_t = \frac{1}{2} \cdot x_{t-1}$
- AR(2): $\hat{x}_t = \frac{6}{10} \cdot x_{t-1} - \frac{3}{10} \cdot x_{t-2}$
- EGRV: $\hat{x}_t = x_{t-s} + s$, s: saison
- $AIC = 2k - 2 \ln(L)$ with $k = \text{\#parameters}$ and $\ln(L) = - \sum_{t=k}^T \epsilon_t^2$, ϵ = residuals (error)
(the smaller, the better)